OTHER-REGARDING PREFERENCES

AND

SOCIAL NORMS

Lynn A. Stout*

March 2, 2001

* Lynn Stout is a Professor of Law at the Georgetown University Law Center and the Director of the Georgetown-Sloan Project on Business Institutions. Research on this article has been supported in part by the Alfred P. Sloan Foundation through its funding of the Georgetown-Sloan Project, and by the Washington University School of Law Center for Interdisciplinary Studies.

I. INTRODUCTION

Several months ago, I returned home after a rather lengthy out-of-town business trip. I waited for a taxi at the airport, got in, and gave the driver my address. By the time the taxi arrived at my home it was dark. The fare was $16, and I fumbled in my wallet and pulled out what I thought was a $20 bill. I gave to the driver and told him to keep the change. There came a moment of silence. Then the driver said, "You just gave me a $50 bill."

In the early 1980s, when I was taking graduate level courses in economics and in law, my instructors taught me that events like this should not occur. And if they did occur, I should not to pay attention to them, because such behavior was both uncommon and unpredictable. I was taught that the best way to model human behavior was to assume that people always behaved like _homo economicus_ -- that perfectly rational, perfectly selfish actor. Both I and my instructors knew, of course, that real people did not always behave this way. But departures from rational selfishness were presumed to be rare and capricious, and not worth trying to consider.

Times change, and these days even scholars who are sympathetic to rational choice -- and I fit myself into that category -- have begun to question the wisdom of always assuming that people behave in a rationally selfish manner. This trend is especially obvious in the legal literature. In just the last few years, legal scholars have published a flurry of articles investigating how human choices are distorted by overconfidence, framing effects, anchoring effects, availability biases, and similar deficiencies of human cognition.

There is a curious imbalance to this new "behavioral law and economics" literature, however. In brief, contemporary challenges to the rational selfishness model of human behavior tend to focus far more on the first adjective -- the assumption of rationality -- than on the second -- the assumption of selfishness. In this essay I reverse that emphasis. Instead of focusing on the myriad ways in which human beings act illogically, I focus on the many ways in which we act as if we care about the costs borne and the benefits enjoyed by others.

I shall refer to this phenomenon as _other-regarding preferences_. I want to emphasize that in adopting this phrase, I am employing the word "preferences" in its most narrow and technical economic sense. In other words, I am describing behavior, rather than motivation. I make no attempt to determine what might subjectively inspire one person to look out for another's interests. Pride, guilt, or love may be responsible, or something else entirely. The point is that people sometimes _do_ behave as if they care...
about costs and benefits to others. And when they do, in the economist's parlance they have "revealed a preference" for taking account of others' welfare.\(^3\)

I want to focus on the phenomenon of other-regarding preferences for two reasons. First, I am a bit of an optimist. To an optimist, the currently popular task of cataloging people's various cognitive deficiencies can be a bit depressing, as it often leads to the conclusion that someone who is left to make her own choices will use this freedom to shoot herself in the foot. The natural implication is that people are flighty, neurotic and weak-minded, and these deficiencies must be either remedied or compensated for if we are to maximize social welfare.

In contrast, the phenomenon of other-regarding preferences casts a much more flattering light on human nature. This is because other-regarding preferences have tremendous potential to make people better off. As an economist would put it, other-regarding preferences can be highly efficient. The reason has to do with the fact that one of the fundamental problems that economic theory wrestles with is the problem of inefficient "externalities" -- people's failure to take account of the costs they impose on others or the benefits they withhold from them. Other-regarding preferences can "internalize" what would otherwise be external costs and benefits. As a result, other-regarding preferences can permit people to achieve much better outcomes -- in terms of both their individual and aggregate social welfare -- than they could achieve through selfishness.

I have a second reason for wanting to draw attention to the phenomenon of other-regarding preferences, however. In brief, other-regarding preferences may be far more common and important than is generally recognized. Extensive empirical evidence suggests that people often behave as if they are keeping at least one eye on the welfare of others. Moreover, far from being rare or mercurial, this tendency towards other-regarding behavior is pervasive, powerful, and to a great degree predictable. That observation in turn suggests that a solid understanding of the phenomenon of other-regarding preferences is essential to a solid understanding of the behavior of both individuals and societies.

In illustration, I explore here the role that other-regarding preferences may play in creating and enforcing social norms. As I noted earlier, legal academics recently have become intrigued by behavioral approaches to the law. Yet this fascination may by outstripped by their preoccupation with a second emerging form of scholarship -- the scholarship of social norms. Over the past decade, law journals have published a host of individual articles and symposia on the topic of how norms can regulate behavior and how they interact with legal rules in the process.\(^4\) Even a cursory review of this literature quickly reveals a puzzle, however: there appears to be remarkable uncertainty among legal scholars about how norms work and even what norms are. A general consensus holds that norms are rules of behavior that people follow for some reason other than the fear of legal sanction. Beyond this area of agreement, norms scholars disagree significantly in their views of how norms should be defined and why people follow norms-based rules.\(^5\)

In this essay I address the norms puzzle by suggesting that there may be an important linkage between these two emerging areas of scholarship (behavioral analysis of law and the study of social norms). In particular, I explore how the phenomenon of other-regarding preferences may offer important insights into what norms are, how norms work, and what sorts of norms are most likely to emerge under what circumstances. But before I consider how other-regarding preferences shed light on social norms, I must first address an antecedent question: do other-regarding preferences exist?

**II. SOME EMPIRICAL EVIDENCE ON OTHER-REGARDING PREFERENCES**

Everyday life is full of anecdotal evidence of other-regarding behavior. Travelers leave tips in
restaurants they do not expect to visit again; strangers on the street offer directions to the lost; cab drivers even stop passengers from paying mistakenly high tips. Yet such anecdotal evidence, alone, may not be enough to convince the dedicated skeptic that people are not purely selfish. The reason has to do with the fact that contemporary life is usually arranged so that behavior that shows consideration for others also is at least consistent with (if perhaps not fully explained by) observable external incentives.

Suppose, for example, that I walk down the street and no one mugs me. (I have performed this experiment successfully on many occasions and in many locations.) One possible explanation is that most people would prefer not to harm me just to get the contents of my wallet. However, it is possible that I am not mugged because some people fear that if they did this, they would be arrested and thrown into jail. It is also possible that would-be muggers fear that if they threatened me, I would retaliate and harm them. (Unlikely, but not entirely out of the question). Finally, perhaps some of my fellow pedestrians are deterred from mugging me by the concern that they might be observed by others who know them, and who would carry news of their misbehavior back to their own neighborhoods and sully their reputations.

In day to day life, it is hard to rule out entirely such external forces as reasons for good behavior. For example, I once recounted the story of the cab driver to a colleague. He suggested -- with a straight face -- that perhaps the cab driver told me I had given him a $50 bill because the driver suspected me of being a plainclothes officer sent out by the taxi-regulating authorities. Although this seemed quite unlikely to me, I could not disprove his hypothesis. To convince such a cynic one needs more than anecdotal evidence.

Luckily, there is much more than anecdotal evidence available. A full review of the social sciences literature on other-regarding behavior lies well beyond the scope of this essay. (Indeed, it might make a nice multi-volume treatise.) Instead, I will focus on a particular kind of evidence that ought to persuade even the dedicated cynic. This is the evidence that has emerged from several decades of experiments in which social scientists have placed human subjects into situations quite consciously designed to make their self-interest (as measured by their external rewards and punishments) conflict with the interests of others.

In particular, I will focus on three broad categories of experiments commonly known as social dilemma games, ultimatum games, and dictator games. The results of these experiments conclusively establish that other-regarding preferences exist. They also offer considerable insight when and why other-regarding preferences appear.

A. Social Dilemma Games and Altruism

Let us consider first the social dilemma experiment. As its name suggests, the social dilemma experiment is based on the familiar prisoner's dilemma of game theory. However, where the archetypal prisoner's dilemma involves only two people, social dilemma games can be played by more (sometimes quite a few more) than two players. As in a prisoner's dilemma, each player in is given the option between choosing a "cooperative" strategy or choosing instead to "defect." As in the prisoner's dilemma, an individual player always does best (that is, maximizes her personal payoff) by defecting, no matter what the other players choose to do. As in the case of the prisoner's dilemma, however, the group gets the greatest aggregate payoff if all its members cooperate. Social dilemmas thus are structured so that no rationally selfish player would ever choose to cooperate. Yet this relentless pursuit of self-interest ultimately leaves both the group and its individual members worse off.
Over the past four decades, social scientists have reported the results of countless studies designed to test how real people behave when asked to play a single social dilemma game with strangers. Rational choice analysis predicts a zero percent probability that a subject in such a "one-shot" game choose to cooperate. In other words, defection is always the dominant strategy for *homo economicus*. Yet when *homo sapiens* play a social dilemma game, experimenters observe cooperation rates averaging about 50 percent. What does this finding tell us? Most obviously, that other-regarding preferences exist and indeed are common. A subject in a social dilemma who cooperates is choosing an option that quite plainly serves the group's interest more than her own. There are several lay terms available to describe this type of other-regarding preference, including *kindness, consideration, generosity, sympathy*, and (more generally) *altruism.*

**B. Ultimatum Games and Spite**

Altruism is not the only type of other-regarding preference that has been observed in experimental subjects, however. Consider a second type of game, the *ultimatum game*. The typical ultimatum game involves two players. The first player (called the "proposer") is given a stake of money -- say, $10. The proposer is then told that she can offer to give any portion of it she chooses -- all, a lot, a little, or nothing -- to the second player. The second player (called the "responder") then has a choice of his own. He can choose to accept the proposer's offer. In this case the $10 will be divided as the proposer suggests. The responder can also choose to reject the proposer's offer. If the responder rejects the offer, both players get nothing.

It is clear what *homo economicus* would do in an ultimatum game. The proposer would offer the minimum amount of money possible short of offering nothing (say, one penny) and the responder would accept this minimal amount. It is also clear from the experimental results that real people don't play ultimatum games this way. When people play ultimatum games, the proposer usually offers the responder a substantial portion of the stake, often half. And -- this is even more interesting -- if the proposer does not do this, the responder frequently responds by rejecting the offer.

Revenge is sweet. But in an ultimatum game, it is not costless. When a responder in an ultimatum game rejects any positive offer, he loses an opportunity to make himself better off (as measured by external payoffs) than he was before. Why does he choose to do this? Possibly to punish the proposer. If the altruism seen in social dilemma games is the light side of other-regarding preferences, ultimatum games give us a glimpse of the dark side. Responders who reject offers that they perceive are be "too low" are displaying are willingness to incur a personal sacrifice not to help another but to *harm* her. Synonyms for this sort of other-regarding behavior include *malevolence, antipathy, vengefulness, and spite.*

**C. Dictator Games and Second-Order Effects**

Ultimatum games accordingly demonstrate that other-regarding preferences can take the form of a willingness to sacrifice to harm others, as well as a willingness to sacrifice to help them. But ultimatum games teach us something else as well. To see what, we need to compare the behavior observed in ultimatum games with that observed in a similar but slightly different sort of game: the *dictator game*.

There are two players in the typical dictator game. The first player, like the proposer in an ultimatum game, is given a stake of money and invited to propose a distribution rule for dividing that money between herself and the second player. However, a dictator game differs from an ultimatum game in an
important respect: in a dictator game, the second player is not given any choice or any right to veto the
first player's division of the loot. This is why the first player is now called a "dictator." The second
player gets what the dictator is willing to give up, no more and no less.

Most people who play the role of the dictator in a dictator game choose to give the other player at least
some portion of their initial stake. In other words, most dictators display at least some degree of
altruism. However, while dictators in dictator games usually share their wealth, on average they do not
share as much as proposers in ultimatum games do. Offers in dictator games tend to be smaller than
offers in ultimatum games. This suggests that in addition to altruism, proposers in ultimatum games
may have a second motivation for their willingness to share their stakes -- the fear that the responder
might react to a low offer by spitefully rejecting it.

In other words, proposers in ultimatum games fear responders' vengeance. This is a very interesting
finding, for it suggests not only that people may have other-regarding preferences, but that they know
that other people may have other-regarding preferences. That possibility in turn suggests that other-
regarding preferences influence human behavior at least two levels.

At the first level, other-regarding preferences can cause some people to sacrifice to either help or to
harm others around them. In other words, people who have other-regarding preferences will behave
differently than they would if they were purely selfish.

At the second level, the knowledge that some people have other-regarding preferences will lead other
people to alter their behavior in reliance upon this possibility -- even if those others are themselves are
purely selfish. For example, suppose Mary is selfish. She still might choose to deliberately make herself
vulnerable to John if she expects John to behave altruistically. This sort of reliance behavior might be
called "rational trust." Similarly, rationally selfish Mary might avoid taking advantage of John's
vulnerability if she believes that if she were to take advantage, John would be willing to incur a cost to
spitefully punish her. This might be called "rational fear of vengeance."

D. Determinants of Other-Regarding Preferences in Experimental Games

At this point it is possible to summarize at least three important lessons to be learned from social
dilemma games, ultimatum games and dictator games. In brief, other-regarding preferences exist; they
come in both positive and negative flavors; and they have both first- and second-order effects on human
behavior.

As interesting as those observations may be, they are not yet very useful. To predict human behavior --
more, to influence human behavior -- we need to know other things as well. Most obviously, we need to
know when people are likely to display other-regarding preferences.

Let us return again to the typical 50 percent cooperation rate observed in many social dilemma
experiments. This result supports the claim that people can behave in an other-regarding fashion. But
it also supports the claim that people can behave selfishly. (After all, if people were purely altruistic, we
would observe 100 percent cooperation rates). What can explain why people some people cooperate and
others don't, or why the same person may cooperate at one time and not at another? What determines
when we are altruistic, when we are spiteful, and when, like Rhett Butler, we don't give a damn?

Answers to these questions are available. Over the past four decades, social scientists have published the
results of literally hundreds of studies of human behavior in social dilemmas, ultimatum games, and
dictator games. These studies have produced a wealth of evidence on who is likely to behave in an other-regarding fashion and under what circumstances they are most likely to do this. Taken as a whole, the evidence strongly supports the following proposition: whether or not people behave in an other-regarding fashion is determined largely by social context, tempered -- but only tempered -- by considerations of personal cost.

1. The Role of Social Variables

Perhaps the most consistent and striking finding that has emerged from the experimental studies is that the behavior of subjects in social dilemma, ultimatum and dictator games appear to be driven largely by social variables. In other words, the subjects' decisions whether or not to behave in an other-regarding fashion appear largely determined by their perceptions of how their behavior will affect others; what others expect and desire of them; how others are themselves likely to behave; and the nature of their relationships with others.

To give a sense of how exquisitely sensitive people are to such social cues, recall the 50 percent average cooperation rate commonly observed in one-shot social dilemmas. This average obscures the fact that cooperation rates in social dilemma experiments can vary widely, and appear to be highly manipulable. By altering particular variables, experimenters have been able to reliably elicit cooperation rates ranging from a low of 5 percent to more than 95 percent. (To appreciate what a truly astonishing degree of range this is, simply recall that the payoff function in a social dilemma is structured so that a rationally selfish player would always defect.) What sorts of social variables have proven important in determining cooperation rates in social dilemma games? Researchers have found that cooperation rates can be raised by allowing the players to speak with each other; by promoting or discouraging a sense of group identity among the players; and by increasing the vulnerability of one's fellow players (that is, the size of the loss to the group if a self-interested strategy is chosen). In this essay, however, I want to highlight two social variables that appear to be especially important (in a statistical sense) in determining other-regarding behavior in experimental games. These are: (1) instructions from authority; and (2) whether subjects believe that their fellow players will behave in an other-regarding fashion.

Studies have consistently found that subjects in a social dilemma games tend to do what the experimenter instructs them to do. If the experimenter says "cooperate", they cooperate; if the experimenter says "defect," they defect. This behavior is puzzling from a rational choice perspective, because instructions do not alter the objective payoffs in the game. Nevertheless, people are so sensitive to directions from authority that they will change their behavior in response to mere hints about what the experimenter desires. For example, in one experiment, subjects playing a social dilemma were told that they were playing the "Community Game." The experimenters observed a cooperation rate of 60 percent. When a group of similar subjects was told they were playing the "Wall Street Game," the cooperation rate dropped to 30 percent.

Similarly, a second social variable that appears to have a strong impact on other-regarding behavior in experimental games is whether a player perceives the other players as other-regarding. This is perhaps most obvious in the case of the ultimatum game, where the perception that a proponent is too "selfish" (self-regarding) often triggers other-regarding behavior in the form of spite from the responder. Similarly, social dilemma studies have found that the expectation of altruism triggers altruism: subjects who believe that their fellow players are likely to cooperate, are far more likely to cooperate themselves. This second result is particularly good evidence of how social factors can overshadow economic concerns in determining other-regarding behavior in social dilemma games, because increasing the probability that other players will cooperate actually increases the economic returns from
defecting.

2. The Role of Economic Variables

In emphasizing the critical role social cues play in eliciting other-regarding behavior, I do not mean to imply that economic payoffs are irrelevant. To the contrary, a second significant finding that emerges from a review of the experimental literature is that other-regarding preferences depend not only on social context but on personal economic payoffs. It is important to recognize that this is not the same thing as saying that people are rationally selfish: any degree of cooperation in a social dilemma, and any significant sharing in an ultimatum or dictator game, is inconsistent with the *homo economicus* model. But the observation that people are capable of both benevolence and malevolence does not mean that they are indifferent to costs associated with these behaviors. When people indulge in altruism or spite, they appear to keep one eye on self-interest in doing so.

This phenomenon is perhaps most clearly observable in social dilemma games, where studies have found that as the personal cost associated with cooperating in a social dilemma rises (that is, as the expected gains from defecting increase), cooperation rates tend to decline. Similarly, if a proposer offers a relatively larger share in a dictator game, the likelihood that the responder will spitefully reject it decreases. Such results suggest that the supply of other-regarding behavior is, in a sense, "downward-sloping." The higher the personal cost associated with other-regarding behavior, the less likely a person will indulge in it. In other words, people are most likely to be are altruistic when altruism is inexpensive, and most likely to be spiteful when spite is cheap.

E. On the Efficiency of a Capacity for Other-Regarding Preferences

The observation that other-regarding behavior depends on both social context and personal cost leads to an interesting conclusion. Most people seem to have at least two personalities or preference functions. The first might be described as a "self regarding" personality. When this personality or preference function is dominant, an individual will seek to maximize her personal payoffs without any apparent regard to how her behavior may affect others. Most people have a second, more "other-regarding" personality, however. When the other-regarding personality dominates, a person will take into account of how her conduct affects others, at least to some extent. In other words, like the Roman god Janus, we all wear two faces. Which we choose to present to the world in any particular situation is determined largely by social context, tempered by considerations of personal cost.

From a purely intellectual standpoint, this is surely an intriguing perspective on human nature. But it may also be of tremendous practical consequence. This is because our human capacity to adopt other-regarding preferences may be of vital importance in promoting the welfare of both individuals and societies.

To understand why, take a moment to imagine a world in which everyone was in fact perfectly selfish. Not to put too fine a point on it, *homo economicus* is a sociopath. She plots her course through life with complete indifference to the wake she leaves behind, including both injuries she inflicts on others and blessings she bestows. Unencumbered by pity or remorse, she will lie, cheat, steal, break promises -- even murder -- whenever a cold calculation of the cost and benefits leads her to conclude that doing so will make her better off. Put in economic terms, *homo economicus* is utterly indifferent to the external costs and benefits that flow from her actions.

Of course, there are a variety of tools available to internalize externalities. These include the legal sanctions imposed by a coercive state; voluntary exchanges in the market; the threat of retaliation in
repeated dealings, and, in smaller communities where people know (or at least know of) each other, reputational concerns. But are such external forces, alone, enough to always control the depredations of homo economicus? The answer to this question must be "no". And this is especially true in developed economies characterized by large populations, free migration, complex production, specialized investment, anonymous exchange, and a high degree of uncertainty -- in other words, in economies like our own.

The reason has to do with the fact that external incentives, alone, can only influence the behavior of the rationally selfish actor when two criteria are met. First, her behavior must be observable to others. Second, some one (or something) must be both willing and able to reward her good behavior and to punish her bad behavior -- and to reward or punish sufficiently.

Even a moment's reflection will quickly suggest the myriad circumstances in modern life where one or another criterion won't be met. On any given day, the average person will be presented with a number of opportunities to impose external costs without fear of detection or effective punishment (e.g., littering, shoplifting, breaking promises, jumping queues, shirking at work). With a bit of forethought, a purely selfish person could identify many more -- and more profitable -- opportunities (burglary, fraud, contract breach, the manufacture of illegal drugs, murder for hire). The net result is that external incentives, alone, often will not be enough to motivate a purely selfish actor to fully take account of the external costs and benefits of her conduct.

So we come to the role of other-regarding preferences. To the extent that people adopt an other-regarding preference function, they will be motivated to benefit others even when they receive no obvious reward for doing so, and to refrain from harming others even when they would suffer no likely punishment. Other-regarding preferences "internalize" externalities -- without any need for time-consuming contract negotiation and drafting, for expensive monitoring, or for costly enforcement measures. In short, other-regarding preferences are often efficient.

This may explain why we have evolved a capacity for them. Evolutionary psychologists have long argued that a capacity for altruism can be highly adaptive in social species.\(^{(29)}\) This is because cooperation that may be "irrational" from the perspective of the individual member of the group can often enhance the group's overall welfare. And if the group does well, individual organisms within the group also do well. For similar reasons, other-regarding preferences may play an evolutionary role not only in the evolution of social species, but in the evolution of social institutions. This is because institutions that promote cooperation among members of a particular social group can allow that group to survive and thrive at the expense of other groups whose institutions do not encourage cooperative behavior.

Law, of course, is a social institution. I have written elsewhere on how the experimental evidence on other-regarding preferences may offer insights into a variety of important questions about how and why law works.\(^{(30)}\) In this essay, however, I would like to focus on how other-regarding preferences may shed light on a second form of social institution -- social norms.

**III. OTHER-REGARDING PREFERENCES AND THE LITERATURE OF "LAW AND SOCIAL NORMS"**

Sociologists have long incorporated the concept of norms into their analyses of human behavior. More recently, however, the idea of norms has come to the attention of the legal academy as well.\(^{(31)}\) In the past decade, a number of prominent legal scholars -- including but not limited to Lisa Bernstein, Robert Cooter, Robert Ellickson, Richard McAdams, Eric Posner, and Cass Sunstein -- have published articles
Legal scholars who write on the topic of social norms often disagree in their exact definitions of what social norms are. There seems to be a general consensus, however, that norms are rules of behavior that are enforced primarily not by courts, but by other forces. For example, in the United States there are norms prescribing that one ought to wear shoes in the office, and to ask permission before smoking in another's house. People perceive such norms as serious constraints on day to day behavior. This is true even though it is extremely unlikely that a decision to take off one's shoes at work, or to "light up" at a dinner party, would trigger a criminal investigation or a civil suit for damages.

Why, then, do people follow norms? It is not unusual for legal scholars who write on the subject of norms to propose that under certain circumstances a norm might become "internalized," so that people obey it even when they would suffer no adverse consequences if they did not. There is an interesting pattern to the way legal scholars tend to talk about social norms, however. In brief, many if not most rely heavily on the idea that norms are followed not primarily because they are internalized, but because someone who violates a norm can expect to suffer a range of external but nonlegal sanctions, including loss of reputation as well as raised eyebrows, disparaging remarks, and other social "punishments." In other words, much of the "new" norms scholarship continues to rely, implicitly or explicitly, on the *homo economicus* model of human behavior.

I believe that the emerging literature on law and social norms is important, and offers a variety of useful insights to legal scholars. I also believe, however, that we can understand how and why norms work far better if we adopt a model of human behavior that acknowledges and incorporates the reality of socially contingent, other-regarding preferences. To illustrate, I briefly consider below how the empirical phenomenon of other-regarding preferences sheds light on three important questions that have been raised in the norms literature. These questions are (1) what sorts of norms are most likely to emerge in a society; (2) why are norms followed; and (3) how can policymakers and other norms "entrepreneurs" deliberately employ social norms to change human behavior.

**A. Other-Regarding Preferences and the Evolution of Norms**

One of the most interesting questions posed by the idea of social norms is the question of why particular norms emerge in particular societies. A related question is why some sorts of norms are more likely than others to prove "sticky" (i.e., more likely to be internalized and followed even when external incentives are missing). These questions are, of course, of critical importance in understanding the value of norms as constraints on human behavior. After all, if norms evolve randomly -- that is, if any form of behavior is just as likely to become and to remain a norm as any other -- why should we conclude that norms provide a social benefit? Indeed, isn't it just as likely that they could impose social costs?

The phenomenon of other-regarding preferences gives us reason to suspect that norms do not evolve randomly, however. To understand why, let us stop to consider some of the norms we observe in our own society. Many of these norms do indeed have a rather arbitrary and fatuous quality (e.g., the norm that adult males ought to wear ties, or that lawn grass ought to be mowed). Norms like these tend to vary from time to time and from place to place. But there is another subcategory of norms that is made up of norms that are far more important and more universal, and that are seen many developed societies. Examples include the norm of abiding by the law even when the police are absent; the norm of keeping one's commitments to others; and, more generally, the norm of treating others as you would like them to treat you.
It should be obvious what these "deeper" norms have in common. They are all significantly other-regarding. The behavioral literature offers an explanation for this tendency. Other-regarding behaviors are especially likely to be socially "codified" into norms, and especially likely to prove sticky, because human beings are predisposed to adopt other-regarding preference functions. Thus an employee who is working late at the office one night might readily choose to violate a dress norm by loosening his tie or taking off his shoes. He is far less likely, however, to help himself to the office supplies.

The notion that people may be particularly inclined to adopt, and to then internalize, other-regarding norms sheds light on an important and ongoing debate in the norms literature: the debate over whether we should expect norms to favor efficient or inefficient behaviors. As the necktie demonstrates, inefficient norms can develop and persist. Nevertheless, a number of scholars have argued that efficient norms are more likely to survive, and that as a result norms tend to evolve in the long run to favor efficient behaviors. The experimental evidence on other-regarding preferences lends support to this view by suggesting why "sticky," other-regarding norms are more likely to have staying power.

It also suggests why other-regarding norms may be more likely to arise in the first place. Robert Ellickson has argued that norms tend to support efficient behaviors because people have an innate preference for utilitarian (welfare-improving) norms. The experimental evidence on human behavior in social dilemma games provides strong empirical support for Ellickson's view. When the social conditions are right -- when an authority tells us we ought to look out for others, when we believe those others are also other-regarding -- we in fact behave like intuitive utilitarians. That observation in turn suggests that people are likely to recognize and prefer social norms that promote utilitarian behavior.

B. Other-Regarding Preferences and the Enforcement of Social Norms

Let us now consider the question of why people obey norms. As the discussion above suggests, one reason why people may obey other-regarding norms is because they "internalize" them (at least when the social and economic conditions are favorable). The experimental evidence on other-regarding behavior has more interesting insights to offer, however, into why people obey norms. In particular, it sheds light on the otherwise-puzzling phenomenon of third-party norms enforcement.

Contemporary norm scholars frequently emphasize the role that bystanders play in enforcing norms by punishing norm-breakers through social sanctions such as dirty looks, disparaging remarks, ostracism, and the like. I suspect this emphasis springs from a desire to keep the study of norms within the parameters of rational choice analysis. After all, if bystanders commonly employ social sanctions against norms-breakers, even homo economicus might be tempted to adhere to norms of behavior in order to avoid such punishments.

Closer inspection reveals, however, that the idea of third-party enforcement poses some problems for rational choice analysis. For example, one can see why homo economicus would want to avoid social sanctions that take the form of others refusing to do business with her. But why should she care if her behavior provokes raised eyebrows? Even more troubling is the notion that third parties are willing to expend significant personal resources in order to enforce norms. After all, rude stares and disparaging remarks can provoke conflict. Conflict is costly. Why should a third-party norms enforcer be willing to incur that cost, when the benefits of her vigilance are shared by the larger society?

Norms scholars have offered some potential explanations. For example, Richard McAdams has argued that people value the esteem of others, and that third party norm enforcers can bestow or withhold esteem at no personal cost. Similarly, Eric Posner has suggested that third-party norms enforcers are
motivated by the desire to signal to potential exchange partners that they are trustworthy, "good types" to deal with.\(^{44}\) It should be noted that both explanations seem in tension with rational choice. Why would a purely selfish person care about others' opinions? And why would someone whose sole motivation is to impress others ever enforce a norm absent an audience of potential exchange partners?\(^{45}\)

The phenomenon of other-regarding preferences suggests another, potentially simpler explanation for third-party norm enforcement. In brief, this behavior reflects other-regarding preferences in the form of spite. The argument goes as follows. As noted earlier, other-regarding behavior often allows groups of individuals to achieve better average outcomes than they could through rational selfishness.\(^{46}\) Thus people who obey other-regarding norms may reasonably perceive individuals who violate these norms as self-regarding recalcitrants whose noncooperation threatens the welfare of the group. This perceived threat triggers the adoption of other-regarding preferences that make them willing to incur the personal cost associated with punishing norms-breakers through dirty looks, disparaging remarks, and confrontations.

Such vengeful "shaming" may not be pretty to watch. But it is, quite possibly, efficient. By increasing the cost of selfish behavior, social sanctions decrease the relative cost of obeying norms of altruism. And as we have seen, when the cost of altruism declines, the amount supplied increases.\(^{47}\) Thus third-party norms enforcers serve the group's welfare by keeping would-be sinners from straying.

C. Other-Regarding Preferences and the Manipulation of Social Norms

So we turn to the third and final question: what does the behavioral evidence on other-regarding preferences tell policymakers and other "norms entrepreneurs" about how they can deliberately employ norms to control behavior?

Perhaps the most obvious and important lesson to be drawn from the evidence is that social context matters. People can be motivated to adopt other-regarding norms; to follow norms even when they have no external incentives to do so; and to enforce norms against others even when this is personally costly. But they can only be motivated to do these things when the social conditions are favorable.

When are the conditions "favorable"? The experimental evidence reviewed in Part II suggests some obvious possibilities. As noted earlier, subjects in experimental games are much more likely to behave in an other-regarding fashion if the experimenter indicates tells them they should do this.\(^{48}\) This intimates that other-regarding norms are more likely to be adopted and followed when people believe they enjoy the support of a respected authority. Courts and legislatures, obviously, can play the role of such an authority. The natural implication is that courts and legislatures can change or support norms through their pronouncements of what people "ought" to do, and so influence behavior without actually imposing legal sanctions. In other words (as many scholars have suggested) the law can change behavior through its "expressive function."\(^{49}\) Conversely, when a social norm is not supported by the law (or by some other respected authority, such as a religious institution), it will likely prove far more ephemeral.

Similarly, if we want people to conform to other-regarding norms, the experimental evidence suggests that it is important to promote a perception that others in society are also conforming to those norms.\(^{50}\) This observation lends additional support to the "broken windows" school of law enforcement, which posits that when individuals in society observe evidence that others are behaving in a criminal fashion, they are more likely to commit crimes themselves.\(^{51}\) Commentators who support the broken windows
thesis sometimes argue that visible signs of disorder encourage crime because they send a signal that the police force is either unwilling or unable to police effectively, which makes criminality seem more appealing by reducing the apparent probability of apprehension and punishment. The experimental evidence on other-regarding behavior suggests a second explanation: visible crime encourages self-regarding behavior because it signals that others are being selfish. Conversely, focusing enforcement efforts on such highly-visible but minor "quality of life" crimes as loitering or defacing property may be an effective strategy for deterring more serious crimes, because it encourages altruism by sending the social message that others are behaving in an altruistic fashion.

It is important for the norm entrepreneur to bear in mind, however, that while social context may be vitally important, it is not the sole determinant of other-regarding behavior. Personal cost is a factor too. Thus social norms may work best at promoting other-regarding behavior when obeying (or enforcing) the norm is not too burdensome. Put differently, norms may be a highly efficient means of inducing individuals to make small contributions to others' welfare -- small contributions that over time add up to large aggregate effects. But we cannot ask norms to bear too much weight. When compliance with other-regarding norms becomes too costly, compliance falters.

IV. CONCLUSION

Rational choice analysis offers a wide variety of important insights into the behavior of both individuals and societies. At the same time, rational choice falls short in its attempt to explain a number of patterns of conduct we observe in daily life. One such pattern is adherence to social norms.

In this essay I argue that we can make far more sense of the idea of norms if we abandon the homo economicus model in favor of a behavioral model that recognizes an important reality of human nature: the reality of socially contingent, other-regarding preferences. Extensive empirical evidence supports the claim that most people shift freely between self-regarding and other-regarding modes of behavior, depending on their perceptions of social context and relative personal cost. This phenomenon is neither rare nor capricious. To the contrary, it is endemic and predictable. And as a result, it will often be of vital importance to a sound understanding of many phenomena, including norms.

I would like to close by returning to my case of the other-regarding cab driver. In the months following this incident I have had the occasion to think about it quite a bit. More important, I have had occasion to think about it in light of the extensive evidence social scientists have amassed on human behavior in social dilemma games, ultimatum games, and dictator games. That literature has lead me to two conclusions.

First, I should not have been surprised (or at least as not a surprised as I was) by the driver's behavior. As I observed to him at the time, there are many taxi drivers out there who would not have felt compelled to warn me from mistakenly giving an enormous tip. But there are many who would have.

Second, I have concluded that while I should be grateful to the driver for telling me that I had handed him a $50.00 bill, the driver should not be the sole object of my gratitude. Certainly he is a fine fellow. But I am grateful as well for the social context that promoted his kind and considerate act. I am grateful that some authority at some time -- his boss, his religious leader, his mother -- taught him that it was morally wrong, even in a nakedly commercial transaction, to take advantage of someone else's clear mistake. I also am grateful that I clearly communicated my own intent to give him a good tip. Through this behavior, I signaled my own intent to behave in an other-regarding fashion.

Finally, I am grateful that I didn't give him a $100 bill.
1. Rational choice analysis has both normative and positive aspects. From a normative perspective, rational choice presumes that the goal is to efficiently allocate resources so as to maximize social welfare (usually interpreted either as maximizing aggregate utility or maximizing aggregate wealth). See Lynn A. Stout and David W. Barnes, Law and Economics ___ (1992). In this essay I adopt the normative goal of rational choice while questioning one of its positive assumptions: the assumption that human behavior conforms to the model of rational selfishness.


3. As this discussion suggests, I am not offering a cognitive model of other-regarding behavior. Although such a model would be extremely useful, social scientists are not in agreement about what motivates other-regarding behavior. Instead I adopt a simple behavioral approach: I treat the human actor as a "black box," and compare the inputs that go into the box with the behavior that comes out. As will be seen, even this simple approach offers a variety of useful insights.


5. Ellickson, supra note __, at __.

6. Other-regarding behavior is often apparently consistent with legal incentives because a variety of legal rules are designed to promote other-regarding behavior (e.g., criminal law, contract law, tort law). See infra text accompanying notes __ (suggesting a reason why other-regarding behavior is often theoretically consistent with legal incentives.) Similarly, many of the acts of altruism (and vindictiveness) we observe in daily life occur between people who are acquainted with each other and who operate in the same community (the neighborhood, the workplace, the family). Thus it is difficult to exclude the possibility that apparently other-regarding behavior actually is motivated entirely by concern for future consequences in the form of reciprocity or reputational loss.

7. This is not, by the way, the explanation that the cabdriver offered. When I complimented him on his honesty, he replied "I have to live with myself." Although this statement clearly suggests some internal enforcement mechanism, the point is that is it is internal.

8. A common example of a social dilemma is the "group contribution game." In this game, a group of \( n \) players is assembled. Each player is given an initial "stake" (say, $10), and given the choice of either keeping the cash or contributing it to a common investment pool. The players are told that any money contributed to the pool will be multiplied by some factor that is greater than 1 but less than \( n \). The money will then be distributed equally among all the players, including those who did not contribute. As
a result the best individual strategy is to keep the $10 (plus a pro rata portion of anything that ends up in the common pool). However, the best group (and average) outcome requires universal contribution.


10. Sally, supra note __, at __.

11. Again, when I say that cooperation does not "serve" the player's interest I am not making a claim about what subjectively motivates cooperation. See supra text accompanying notes __. The motivation could be guilt, sympathy, or ego, all of which may lead the player to conclude that she is psychologically better off (happier, less conflicted) if she cooperates. But whatever the internal mechanism, subjects who cooperate in a social dilemma can be said to "reveal a preference for" (to act as if they care about) serving others' welfare.


13. Camerer & Thaler, supra note __, at 210 ("offers of less than 20 percent are frequently rejected").

14. Camerer & Thaler, supra note __, at 213.

15. Id.

16. See supra text accompanying notes __.

17. See authorities cited supra notes __ (surveying literature).

18. Sally, supra note __, at __.

19. Id. at 78.

20. In social dilemmas, if the players are divided into subgroups and then asked to play with members of another subgroup, cooperation rates fall below those observed when the experimenter makes no attempt to foment any group identity. Id. at 78. See also Camerer & Thaler, supra note __, at 213-14 (reporting study finding that the size of the offers made by dictators in dictator games gets smaller as "social distance" between dictator and recipient increases.)

21. Sally, supra note __, at 79.

22. Id. at 78.

results have been observed in dictator games, where dictators make larger offers when they are instructed to "divide" their stakes than when the experimenters use the "language of exchange". See Camerer & Thaler, supra note __, at 213.

24. See supra text accompanying notes __. Interestingly, responders are far more likely to accept a small offer if they believe it is was generated by a computer than if they think it came from a human proposer. Camerer & Thaler, supra note __, at 214-15.

25. See Scott T. Allison & Norbert L. Kerr, Group Correspondence Biases and the Provision of Public Goods, 66 J. Personality & Soc. Psych. 688 (1994) ("[n]umerous studies have reported that individuals are more likely to cooperate when they expect other group members to cooperate than when they expect others to defect"); Toshio Yamagishi, The Structural Goal/ Expectations Theory of Cooperation in Social Dilemmas, 3 Advances in Group Processes 51, 64-65 (1986) (discussing experimental findings that "expectations about other members' behavior is one of the most important individual factors affecting members' decisions in social dilemmas").

26. See, e.g., Sally, supra note __, at 75 (finding in regression analysis of over 100 social dilemma studies that doubling the reward from defecting decreased average cooperation rates by as much as sixteen percent).

27. See supra text accompanying notes __. This is not as clear evidence, as a larger share might also lead the responder to conclude that the proposer is being more "fair," and so spite is not called for. Cf. Camerer & Thaler, supra note __, at 211 (reporting experiment in which playing ultimatum games for much larger stakes did not significantly change behavior).

28. These observations naturally raise the question: if people are only likely to adopt other-regarding preferences when this is not too personally costly, how can other-regarding preferences produce significant social gains? There are two parts to the answer. First, other-regarding behavior that is quite inexpensive to the actor can produce much larger gains to the recipient. (A good example might be when one person refrains from mugging another, as a mugging often provides less benefit to the mugger than the loss it imposes on the muggee in terms of lost time and property and personal injury). Second, even small efficiencies from other-regarding behavior, when added up over many individuals and many social interactions, can produce great aggregate social gains.


30. See Margaret M. Blair and Lynn A. Stout, Trust, Trustworthiness, and the Behavioral Foundations of Corporate Law, 149 U. Pa. L. Rev. __ (2001); see also Lynn A. Stout, Other-Regarding Preferences and the Law (in progress).


33. Ellickson, supra note __, at 5.


35. Ellickson, supra note __, at 5 ("the new norms scholars all hew to a rational-choice model of human behavior").

36. See generally Eric A. Posner, Law, Economics, and Inefficient Norms, 144 U. Pa. L. Rev. 1697 (1996). It should be noted that norms that encourage behavior that would not necessarily be efficient in isolation may nevertheless promote social welfare if they provide a "focal point" for coordinating behavior, and coordination is itself efficient. See generally Richard H. McAdams, A Focal Point Theory of Expressive Law, 86 Va. L. Rev. 1649 (2000).

37. By this I mean that someone who violates the norm imposes a cost on another above and beyond any offense that other may take from the mere knowledge that the norm was not followed.

38. See Posner, Inefficient Norms, supra note __, at 1697-98 (discussing debate).

39. Ellickson, supra note __, at 29.

40. See supra text accompanying notes __ (noting the likelihood of altruism increases as the benefit to beneficiary increases) and __ (noting the likelihood of altruism decreases as the cost to the altruist increases).

41. See, e.g., Bernstein, supra note __, at 116 (discussing importance of reputational concerns in ensuring compliance with industry norms). Cooter, supra note __ at 1665, 1668-69 (suggesting third-party enforcement is important in explaining emergence of norms); Posner, supra note __, at 1699 ("a norm is like a law, except that a private person sanctions the violator of a norm, whereas a state actor sanctions the violator of a law"); Sunstein, supra note __, at 915 (discussing norms as enforced through social sanctions).

42. See Geoffrey P. Miller and Lori S. Singer, Norm Enforcement in a Noncooperative Setting (1999) (unpublished manuscript on file with author) (discussing numerous accounts of instances where third parties employed social sanctions against strangers to enforce handicapped parking rules at significant personal cost, including threat of injury).

43. See McAdams, supra note __.


45. Indeed, even if there were an audience, it is unclear why they would accept the enforcement action as a signal of a good character, because given the nature of homo economicus, no exchange partner ever expert her to remain "a good type" when her self interest dictated otherwise.

46. See supra text accompanying note __.
47. See supra text accompanying note __.

48. See supra text accompanying notes __.


50. See supra text accompanying notes __.


52. See, *e.g.*, id. at 357.